International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Automating Data Privacy Compliance through Filtering Algorithms

[1]Demelyn E. Monzon, PhD., [2]Leandro Avena IV, MSIT

Submitted: 03/02/2024 **Revised**: 11/03/2024 **Accepted**: 17/03/2024

ABSTRACT: Implementing controls to address identified privacy risks in Business Process Outsourcing (BPO) companies presents a significant challenge for Data Privacy Officers. A well-functioning control system directly contributes to the privacy risk rate associated with each identified issue. Failure to implement controls correctly can escalate the level of privacy risk.

The researcher has developed a unique system that is integrated into a Privacy Impact Assessment (PIA) tool. This system, powered by content and collaborative filtering algorithms, takes a collaborative approach to privacy risk management. Based on the risk category of each control and historical data, it recommends controls for reducing or eliminating the risk associated with each project within the organization. A collaborative approach empowers everyone to feel responsible for the outcome, encouraging all stakeholders to actively participate in addressing privacy risks.

This study adopts a combination of developmental and descriptive approaches. Developmental research focuses on systematically designing, developing, and evaluating the recommender system for the PIA tool, ensuring it meets the requirements. Descriptive research, meanwhile, investigates common privacy risks, implementation challenges, strategies employed, and respondents' satisfaction levels with the developed system.

The thoroughness of the research findings is a testament to the potential risks in BPO facilities. Many of these facilities allow unrestricted employee access to data storage areas, leading to potential breaches. Additionally, technical issues with data processing equipment often result in accidental exposure of personal and sensitive information even after disposing of records. These identified risks directly contravene the security protocols mandated by the National Privacy Commission (NPC), which require strict physical security measures for organizations handling personal data. The comprehensive nature of these findings instills confidence in the proposed solutions, reassuring the audience about the effectiveness of the proposed system.

This research introduces a crucial solution-a recommender system integrated within a Privacy Impact Assessment (PIA) tool. This system, powered by Collaborative filtering and content filtering, is designed to effectively address the challenges posed by privacy risks in BPO companies. Its ability to analyze past assessments and suggest controls based on similar situations, as well as categorize the appropriate control type based on risk description, makes it a valuable tool for Data Privacy Officers (DPOs) and top management.

By utilizing collaborative and content-based algorithms, the system not only recommends privacy risk levels and corresponding controls for identified and newly identified risks but also includes the audience in the process. This assists Data Privacy Officers (DPOs) in reducing risk levels by lessening or eliminating the potential harm from privacy breaches and making informed decisions. The system provides recommendations for top management to ensure compliance, fostering a sense of inclusion and shared responsibility in addressing privacy risks.

Keywords: Collaborative Filtering, Content Filtering, Data Privacy Impact Assessment, Business Process Outsourcing, Data Privacy Officers

I. INTRODUCTION

The Business Process Outsourcing (BPO) sector, characterized by ongoing advancements in software development and privacy protection, is a significant player in the technological landscape of the Philippines. In 2017, the Philippines ranked seventh globally in outsourcing destinations, with cities like Manila, Cebu, and Davao prominently in Tholons' Top 100 global outsourcing locations for 2018. Among the populous cities in the Philippines are Sta. Rosa, Bacolod, and Iloilo are evaluated based on digital transformation, talent pool, and service quality. Currently, there are 851 registered BPO companies in the Philippines, with approximately 400 (46.2%) offering computer or IT-related services, along with nine in animation and twenty in medical transcription (Phillabor, 2019).

Implementing the Data Privacy Act of 2012, locally known as Republic Act (RA) 10173, has significantly impacted BPO companies managing employee data nationwide. This legislation aims to safeguard sensitive personal information and mitigate high-risk practices within organizations, emphasizing the importance of privacy rights while promoting information flow for innovation and development. Recognizing the integral role of information and communication technology in national progress, the law underscores the responsibility of both private and public sectors to secure personal data

[1]*Polytechnic University of the Philippines, Quezon City Campus*
demonzon@pup.edu.ph
ORCID: 0000-0002-3337-1735
[2]*Polytechnic University of the Philippines, Quezon City Campus*
lbavena@pup.edu.ph

(National Privacy Commission, 2016). Additionally, guidelines outlined in National Privacy Commission Advisory No. 2017-03 detail the process of Privacy Impact Assessment (PIA) as a means to evaluate and manage privacy impacts on programs, projects, or systems handled by Personal Information Controllers (PICs) or Personal Information Processors (PIPs) (National Privacy Commission, 2017).

BPO companies, which often handle sensitive personal data, are directly impacted by such regulations. Critical challenges faced include:

- assessing privacy risks associated with various services and projects,
- implementing appropriate controls to mitigate identified risks and
- determining the repercussions of maintaining high-risk processes.

Manual risk assessment is susceptible to human error, posing additional data integrity, confidentiality, and availability risks. Compliance with RA 10173, or the Data Privacy Act of 2012, is crucial for businesses to protect individuals' data and minimize potential harm.

To address these challenges, implementing a recommender system for privacy impact assessment proves beneficial for BPO companies. This system utilizes collaborative and content-based algorithms to recommend controlling identified privacy risks. This approach is optimal for businesses aiming to distinguish themselves from their competitors. (Rocco, 2019). Collaborative methods rely on past user-product interactions to offer personalized suggestions, while content-based approaches consider additional subject information. By employing both algorithms, BPO companies can gather recommendations for managing privacy risks identified by the Data Privacy Officer (DPO).

The study focuses on developing a Recommender System that utilizes collaborative and content-based algorithms for Privacy Impact Assessment (PIA). This system aims to recommend controls for identified and newly discovered privacy risks, aiding the DPO in mitigating or eliminating privacy risks and providing insights for top management decision-making.

II. METHODOLOGY

This study combines developmental and descriptive research approaches. The developmental research focuses on designing, developing, and evaluating a recommender system for Privacy Impact Assessment (PIA) utilizing content and collaborative filtering algorithms. The descriptive research has identified common privacy risks and challenges BPO employees face in implementing solutions for storing, processing, and disposing personal data. It also assessed the strategies used by BPOs and the user satisfaction with the developed system.

In data collection, a survey questionnaire was the primary instrument used to collect data on the common problems BPOs face in implementing privacy solutions. Common privacy risks encountered by BPOs related to data storage, processing, and disposal. Strategies used by BPO employees to mitigate privacy risks. User satisfaction with the developed recommender system. The questionnaire used a five-point Likert scale format for responses.

For the recommender system development, the recommender system was built using the programming language PHP (CodeIgniter framework), with libraries such as filters and impulse ml recommender (based on positive reviews). The system operates in a web environment and requires user credentials for access with permission-based controls.

III. RESULTS AND DISCUSSION

There are common privacy risks encountered by BPO employees when storing, processing, and disposing of personal and sensitive personal information. An identified privacy risk is associated with the storage of personal and sensitive personal information in business process outsourcing companies.

The top privacy risks in BPO data management highlight critical risks identified in BPO data storage practices, such as unrestricted physical access where data cabinets and facilities are accessible to all employees, increasing the risk of unauthorized data breaches. (58.30% & 61.70%). Weak shared server security has shared servers across departments lack access restrictions, creating vulnerabilities for unauthorized access. (41.70%). Unreliable data backups in the absence of a concrete data backup process by the infrastructure department pose a risk of permanent data loss in case of incidents. (40.00%). Personal device usage that allows employees to use personal devices to access company accounts (e.g., webmail) introduces potential security risks if the devices are compromised. (50.00%)

These findings emphasize the need for BPOs to implement stricter access controls, secure shared servers, establish robust data backup procedures, and develop policies on using personal devices.

Privacy Risk Assessment in BPO Data Storage

This study identified privacy risks in BPO data storage practices:

- **Unrestricted Access:** A significant majority (**61.70%**) of respondents reported a lack of controls for accessing rooms and facilities, raising concerns about physical security. It was

followed closely by fears of unrestricted access to cabinets and storage containers (**58.30%**).

- **Insecure Shared Servers:** Over 40% of respondents (**41.70%**) indicated the use of shared servers across departments without access restrictions, which creates vulnerabilities for unauthorized access to sensitive data.
- **Unreliable Backups:** Nearly half (**40.00%**) of respondents expressed concerns about the infrastructure department's lack of a concrete data backup process. This poses a significant risk of data loss in case of incidents.
- **Personal Device Usage:** A substantial portion of respondents (**50.00%**) reported that employees can use personal devices to access company accounts like webmail. This practice introduces potential security risks if the devices are compromised.

These findings highlight the need for BPOs to implement stricter access controls for physical locations and data storage systems. Additionally, establishing robust data backup procedures and policies regarding personal device usage is crucial to mitigate privacy risks.

Privacy Risks in Processing Personal Information at BPOs

A survey revealed several concerns about privacy risks associated with handling personal and sensitive data at business process outsourcing (BPO) companies. Here's a breakdown of the key findings:

- **Technical Issues (65%)**: Many respondents (65%) reported consistent problems with outdated or malfunctioning computer hardware and software. These technical issues can lead to data breaches or accidental loss.
- **Unsecure Communication (56.70%)**: Over half (56.70%) of respondents indicated that employee webmail accounts were frequently targeted with spam, phishing attempts, and scams. Robust email security measures are paramount in today's digital landscape, as highlighted by these findings.
- **Malware Concerns (55%)**: Most respondents (55%) were concerned about malware, such as viruses, Trojans, or worms, on company computers. Malware can compromise data security and lead to unauthorized access.
- **Human Error (46.70%)**: Nearly half (46.70%) of respondents reported witnessing human errors during data processing tasks. The identified risks highlight the necessity of robust training programs and established data handling procedures.
- **Employee Dishonesty (35%)**: A concerning percentage (35%) of respondents suspected employees of potentially compromising data due to dishonesty. Strong ethical policies and data access controls are crucial.
- **Weak Credentials (21.70%)**: Over a fifth (21.70%) of respondents highlighted a lack of security for user credentials, making them vulnerable to theft. Implementing multi-factor authentication and strong password policies is essential.

These results underscore the urgent need for more robust data protection measures. Security practices in BPOs are used to minimize privacy risks associated with handling personal and sensitive information.

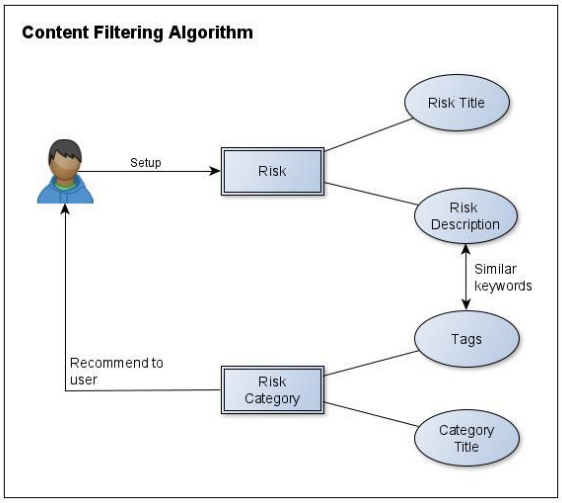

Figure 1: **Content-Based Filtering: A Core Recommender System Technique**

This figure illustrates how the recommender system utilizes a content filtering algorithm to assist users in selecting appropriate risk categories during Privacy Impact Assessments (PIAs). Here's a breakdown of the process:

1. User Input: Users have the flexibility to define risks by providing a title, description, and category.
2. Risk Category Recommendation: When users create a PIA, the recommender system analyzes the risk description. It searches for keywords associated with pre-defined risk categories.
3. Matching and Recommendation: If the system identifies a match between keywords in the risk description and a specific risk category, it automatically suggests that category in a dropdown menu. This recommendation helps users choose the most accurate risk classification for their PIA.

This revised version clarifies the purpose of Figure 1 and explains the steps involved in the content filtering algorithm. It also uses more straightforward language.

Figure 2. Collaborative Filtering Algorithm Application

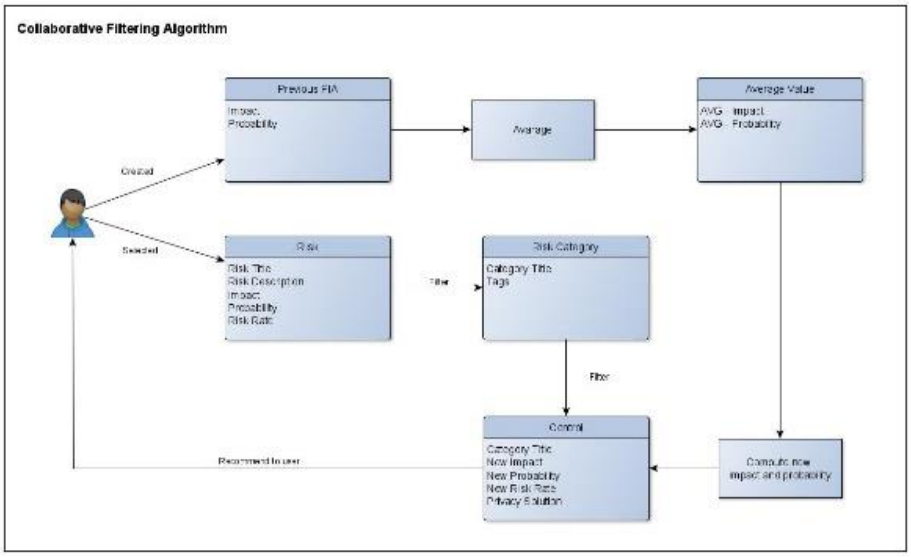

Figure 2: Collaborative Filtering: Recommending Effective Controls

This figure builds upon the concept of content filtering explained in Figure 1. It demonstrates how the recommender system leverages collaborative filtering to suggest suitable controls for PIAs.

1. PIA Creation and Risk Selection: As described earlier, users can create PIAs and define risks with titles, descriptions, and categories (aided by content filtering).
2. Automated Risk Category Population: Once a user selects a risk, the system automatically populates the risk category based on the content filtering algorithm explained in Figure 1.
3. Control Recommendation based on Risk Category: Here's where collaborative filtering comes into play. The system identifies all registered controls with risk categories matching the user's selected risk.
4. Filtering and Recommendation: By filtering the controls based on the matched risk category, the system recommends relevant controls to the user.
5. Impact and Probability Calculation (Optional): While not explicitly mentioned here, the text suggests the system might calculate new impact and probability values. This functionality is unclear and might require further explanation. It's possible this calculation is based on historical data or user input, but the text doesn't specify.

Overall, Figure 2 highlights how the recommender system combines content filtering (explained in Figure 1) with collaborative filtering to provide users with informed control recommendations during the PIA creation process.

This revised version clarifies the connection between Figure 1 and Figure 2, explains the collaborative filtering process, and acknowledges the ambiguity around impact and probability calculation.

IV. CONCLUSIONS

On Privacy Risks in BPO Companies and the Role of PIAs, this research investigates privacy risks associated with personal data handling in Business Process Outsourcing (BPO) companies. The Identified Risks:

- Physical Security Concerns: The study found that most BPO companies lack proper physical

security measures. This includes accessible workspaces, missing dividers, and unsecure storage for equipment, potentially exposing data to unauthorized access.

- Technical Security Issues: The research identified problems with outdated software, hardware malfunctions, and inadequate data disposal practices, all of which increase the risk of data breaches.

The Data Privacy Act (DPA) Requirements: To safeguard personal information, the Data Privacy Act of 2012 mandates organizations to prioritize physical and technical security. This includes securing office layouts, encrypting data, and implementing robust network security. These measures work together to create a strong defense against unauthorized access and data breaches.

The Challenges for BPO Employees: Employees responsible for data privacy face obstacles in implementing solutions due to limitations such as: Lack of physical security barriers (dividers, locks), Reliance on insecure software (free trials, unmaintained systems),Substandard data disposal practices

The Importance of Privacy Impact Assessments (PIAs): Regular PIAs are crucial for identifying and addressing privacy risks. As defined by the National Privacy Commission (NPC), a PIA is a tool that assesses the potential impact on privacy of any process or system that handles personal information. It allows organizations to take necessary actions to mitigate these risks.

Benefits of the Developed PIA Tool: The researcher developed a user-friendly PIA tool accessible on any device with a web browser. It offers several advantages: Simplicity: Easy to learn and use for individuals with varying technical skills, Reliable Results: Provides accurate outcomes based on user-entered data with continuous updates.

Addressing Privacy Risks Through the Tool: This tool incorporates content-based and collaborative filtering algorithms to recommend appropriate controls based on identified risks. These recommendations can help BPO companies reduce or eliminate privacy risks associated with personal data handling.

To conclude, **t**he research highlights the prevalence of privacy risks in BPO companies due to inadequate physical and technical security measures. It emphasizes the importance of PIAs and introduces a user-friendly tool to streamline this process and recommend effective controls for mitigating privacy risks.

REFERENCES

[1] Aghdam, M. (2019). Context- aware recommender systems using hierarchical hidden Markov Model. Physica A: Statistical Mechanics and its Applications, 518(15), 89-98. https://doi.org/10.1016/j.physa.2018.11.037

[2] Alhijawi, and Kilani. (2020). A collaborative filtering recommender system using genetic algorithm. Information Processing and Management, 56(6) https://doi.org/10.1016/j.ipm.2020.102310.

[3] Boström and Filipsson. (2017). Comparison of User Based and Item Based Collaborative Filtering Recommendation Services. Retrieved from http://search.ndltd.org/

[4] Cena, Console and Vernero. (2021). Logical foundations of knowledge-based recommender systems: A unifying spectrum of alternatives. Information Sciences, 546(6), 60-73. https://doi.org/10.1016/j.ins.2020.07.075

[5] Chulyadyo, Rajani. (2016). A new horizon for the recommendation: Integration of spatial dimensions to aid decision making.

[6] Clarke. (2016). Privacy impact assessments as a control mechanism for Australian counter-terrorism initiatives. Computer Law & Security Review, 32(3), 403-418. https://doi.org/10.1016/j.clsr.2016.01.009

[7] Dela Cruz. (2018). Why Do We Undertake Privacy Impact Assessments (PIA)? Retrieved from http://ateneo.edu/udpo/article/Why-do-we-undertake-Privacy-Impact-Assessments-PIA

[8] Disini and Disini Law Office. (2018). Fostering a culture of privacy through the conduct of Privacy Impact Assessments. Retrieved from https://privacy.com.ph/articles/fostering-a-culture-of-privacy-through-the-conduct-of-privacy-impact-assessments/

[9] Dong, Zeng, Koehl, and Zhang. (2020). An interactive knowledge-based recommender system for fashion product design in the big data environment. Information Sciences, 540, 469-488. https://doi.org/10.1016/j.ins.2020.05.094

[10] Eirinaki, Gao, Varlamis, and Tserpes. (2017). Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions. Future Generation Computer Systems, 78(1), 413-418. https://doi.org/10.1016/j.future.2017.09.015

[11] Esmaeili, Mardani, Alireza, and Golpayegani. (2020). A novel tourism recommender system in the context of social commerce. Expert Systems with Applications, 149(1). https://doi.org/10.1016/j.eswa.2020.113301

[12] Gao, Zhang, Yu, Li, Wen, and Xiong. (2021). Recommender systems based on generative adversarial networks: A problem-driven perspective. Information Sciences, 546(6), 1166-1185. https://doi.org/10.1016/j.ins.2020.09.013

[13] Herce-Zelaya, Porcel, Bernabe-Moreno, and Herrera-Viedma. (2020). New technique to alleviate the cold start problem in recommender systems using information from social media and random decision forests. Information Science, 536, 156-170. https://doi.org/10.1016/j.ins.2020.05.071

[14] Sinha and Dhanalaksmi. (2020). Evolution of recommender paradigm optimization over time. Journal of King Saud University – Computer and Information Sciences, 34(4), 1047-1059. https://doi.org/10.1016/j.jksuci.2019.06.008

[15] International Standard Organization (2018) ISO 31000 Risk Management – Principles and Guidelines on Implementation. Retrieved from https://www.iso.org/iso-31000-risk-management.html/

[16] Lovine, Narducci, and Semeraro. (2020). Conversational Recommender Systems and natural language: A study through the ConveRSE framework. Decision Support Systems, 131. https://doi.org/10.1016/j.dss.2020.113250

[17] Madasamy. (2019). Introduction to recommendation systems and How to design Recommendation system. Retrieved from https://madasamy.medium.com/introduction-to-recommendation-systems-and-how-to-design-recommendation-system-that-resembling-the

[18] Margaris, Vassilakis, Spiliotopoulos. (2020). What makes a review a reliable rating in recommender systems? Information Processing and Management, 57(6). https://doi.org/10.1016/j.ipm.2020.102304

[19] Mohamed, Khafagy and Ibrahim. (2019). Recommender Systems Challenges and Solutions Survey. International Conference on Innovative Trends in Computer Engineering (ITCE), 149-155. http://doi.org/10.1109/ITCE.2019.8646645.

[20] Napoles, Grau, and Salgueiro. (2020). Recommender system using Long-term Cognitive Networks. Knowledge-Based Systems, 206(28). https://doi.org/10.1016/j.knosys.2020.106372

[21] National Privacy Commission. (2016). Implementing Rules and Regulation. Retrieved from https://www.privacy.gov.ph/implementing-rules-and-regulations-of-republic-act-no-10173-known-as-the-data-privacy-act-of-2012/

[22] Ojagh, Malek, Saeedi, and Liang. (2020). A location-based orientation-aware recommender system using IoT smart devices and Social Networks. Future Generation Computer Systems, 108, 970-118. https://doi.org/10.1016/j.future.2020.02.041

[23] Pyati and Malawade. (2018). A Study on Risk Assessment Using Probability-Impact Matrix Method for A Multi-Storeyed Residential Building. International Research Journal of Engineering and Technology (IRJET), 05(07), 254-257

[24] Raab. (2020). Information privacy, impact assessment, and the place of ethic. Computer Law & Security Review, 37. https://doi.org/10.1016/j.clsr.2020.105404

[25] Sambhav, Vikesha, Sushama. (2018). An Improved Collaborative Filtering Based Recommender System using Bat Algorithm. Procedia Computer Science, 132, 1795-1803. https://doi.org/10.1016/j.procs.2018.05.155

[26] Scudder, McNevin, Kelty, Walsh, and Robertson. (2017). Forensic DNA phenotyping: Developing a model privacy impact assessment. Forensic Sci Int Genet, 34, 222-230. https://doi.org/ 10.1016/j.fsigen.2018.03.005

[27] Seyyar and Geradts. (2020). Privacy impact assessment in large-scale digital forensic investigations. Forensic Science International Digital Investigation. https://doi.org/ 10.1016/j.fsidi.2020.200906

[28] Su, Zheng, Ai, Shen, Zhang. (2020). Link prediction in recommender systems based on vector similarity. Physica A: Statistical Mechanics and its Applications, 560(15). https://doi.org/10.1016/j.physa.2020.125154

[29] Yassine, Mohamed, and Mohammed. (2021). Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. Simulation Modelling Practice and Theory, 113. https://doi.org/10.1016/j.simpat.2021.102375